# FLASH AND PEEP: A ROBUST METHOD FOR FINDING AND TRACKING DISPLAYS

**M. Schneider[1], A. Butz[2], and A. Krüger[3]**

[1)] German Research Center for Artificial Intelligence DFKI, [2)] University of Munich, [3)] University of Muenster, Germany

## ABSTRACT

We present a two-step method for the localization of displays using cameras and visual markers on screens. Displays can attract attention by flashing in a characteristic pattern and color. Once they're found by a camera, their exact position can be calculated robustly from a set of markers they show. To make the method numerically stable, a certain degree of redundancy must exist in the environment. We describe how this method was implemented in our experimental instrumented environment and show a simple example of the system in operation.

## INTRODUCTION

One way to approach the vision of ubiquitous computing as presented in Weiser (14) is to build instrumented environments, in which rooms, furniture, and everyday objects are either instrumented with computing machinery themselves, or externally augmented by other devices in the environment, such as projectors or head-worn displays. Instrumented environments combine the paradigm of mixed or augmented reality with embedded devices, and thus create a computational and informational layer over the physical world.

Virtually all visions of ubiquitous computing include the notion of space and the concept of situation and location-dependent behavior of devices and services. Augmented reality has a particularly strong focus on the spatial arrangement of information, as it overlays a virtual 3D information space to the physical 3D space. In Butz et al (4) we have presented an approach to structure a user's view at this spatial arrangement of information and to integrate it with the devices used for accessing information in a consistent interaction metaphor.

One basic prerequisite for the correct display of spatially arranged information is, that the devices in the environment know their own position and orientation in space. Hence, spatial tracking and calibration has always been a major issue in AR and MR. Classically, objects are tracked by magnetic, acoustic, or optic methods or combinations of these. Usually, either an external instance, the tracker, determines the position of its tracking targets relative to its own known position (outside-in tracking), or the tracker is attached to the moving device and determines its own position relative to known fiducial markers in the environment (inside-out tracking).

Optical markers or tracking targets are usually either printed patterns on white or reflective material (passive markers) as in Billinghurst (3), or active pieces of electronics, emitting spatial and temporal patterns of light. The core idea of this paper is to use regular displays to display optical markers. In a first step they can blink in a characteristic color and sequence to become easily detectable and make themselves distinguishable for cameras around them. Once their presence is registered by one or more cameras, they can, in a second step, display geometrical patterns of given size and shape, thereby allowing the cameras to determine their relative position.

This scheme nicely corresponds to a two-tier social protocol used in human communication to preserve and focus the scarce resource of attention: If we want to communicate to a person who is currently busy with something else, we first knock, wave or utter an indistinct sound to get her attention, and only then say what we want to communicate. In an instrumented environment containing multiple cameras and displays, displays can follow this scheme in order to ask the environment about their own position. The numerical stability of our approach increases with the number of cameras used to determine a display's position.

After briefly reviewing related approaches, we will show how this method was implemented in our instrumented environment SUPIE[1] and present in a simple example of its operation.

## RELATED WORK

In the AR/MR literature, many tracking solutions have been proposed, which could be used for display tracking in an instrumented environment. Most of them require specialized and expensive hardware (7, 2, 15), which sometimes even has to be attached to every tracked object (9, 1). Since the topic of this paper is a tracking method involving just the cameras and displays already available in the environment, we will only review approaches matching these constraints.

[1] Saarland University Pervasive Instrumented Environment

The concept of matrix-shaped markers and some algorithms for the registration of these markers in images and video streams has first been proposed in Rekimoto (10). Various implementations for different computational platforms exist, allowing even the use of cameras integrated into PDAs (13) and mobile phones (11). The main idea is to use cheap markers printed on paper for marking objects or locations and thereby either identify the position of the marked objects with the camera or inversely derive the position of the camera from known marker positions.

Although this works great for a wide range of applications, some problems prevent the general use in instrumented spaces. The similarity between markers and hence their a priori recognition rate, for example, depend on the overall amount of markers used. On top of this, markers have to possess a minimum size in order to be recognized reliably in the video stream, raising the aesthetic question, whether an environment can be plastered with these markers or not.

The approach presented in Kishino (8) addresses the first restriction by using dynamic markers, either implemented as a special array of LEDs or by displaying dynamically changing grid-based markers on a computer screen. However, cameras still have to be positioned close enough to guarantee a minimum size of the markers in the image. Our proposed two-step method also addresses this second problem, providing an aesthetically superior solution by displaying markers only if needed.

In accordance with our own efforts to build self-configuring environments, Harle (6) present an approach to build a spatial model of the environment with the help of signals emitted by an already installed ultrasonic positioning system. Similarly, approaches from the field of active vision in robotics use controllable camera configurations to efficiently extract landmark features while a robot or human is moving through an unknown environment (5).

## A METHOD FOR DISPLAY LOCALIZATION IN INSTRUMENTED ENVIRONMENTS

The location of a display (or any other object) is mathematically described by its position and orientation within a given coordinate system. Both position and orientation can be derived by optical marker tracking systems, such as the ARToolkit (3).

In this section we will present our idea of a scalable and flexible display tracking system based on an enhanced optical marker concept. Instead of externally attaching printed markers to the displays in question, ”soft markers” as explained in the next section are directly shown by the displays themselves whenever needed. In a two step approach these soft markers are first discovered and then evaluated by an array of remotely controlled cameras.

### Hard and Soft Markers

Although printed optical markers have proven to be a feasible and low-cost solution for the identification and tracking of objects by cameras, their static nature and inflexibility restricts their possible applications in several ways.

To be recognizable within a camera image, markers need to have a certain minimum size, depending on camera characteristics. Bigger markers are easier to recognize, but occupy more space on the tracked object. A tradeoff between tracking performance and marker size has to be found. On some objects, such as mobile phones or PDAs, there will not even be enough space for very small printed markers.

To be reliably distinguishable, different markers have to possess a significantly different optical structure. Similarity between markers will increase with their overall number, which makes it increasingly harder to reliably distinguish them. Therefore, a trade-off between identification accuracy respectively robustness and the number of unique markers has to be found.

Generally, printed markers are assigned to objects at design time and can not be dynamically changed at run time. Therefore, to uniquely identify even a minimal subset of a potentially huge number of objects, a very large set of markers has to be defined. As explained above, this will cause a serious problem in distinguishing different objects even if only a few objects are present.

**Figure 1: Soft marker on a computer screen**

Markers attached to an object are always visible, even if they are not currently needed. Besides purely aesthetical considerations, the presence of many markers will also have a negative impact on the recognition system. Processing and matching of potential marker candidates will consume valuable resources and lead to additional ambiguities, especially when looking out for a certain marker in a large marker ”haystack”.

To gain more flexibility in the application of markers, we introduce the concept of soft markers. These are optical markers, which are not statically attached to an object, but shown on ordinary displays as needed. An example of a set of soft markers applied to a TFT display is shown in figure 1. The name ”soft marker” illustrates both the fact that they are flexible and dynamic, and that they are realized by software components (as opposed to hard markers, which are physically printed onto hardware components). By using soft markers, we can overcome most of the restrictions of printed optical markers mentioned above.

**Figure 2: Pan/Tilt/Zoom camera mounted to the ceiling**

Soft markers can be turned off when they are not needed, both improving the performance of marker recognition and greatly reducing visual clutter. The size of soft markers can be adjusted at runtime, allowing the environment to choose the best compromise between recognition performance and consumed (screen) space depending on the concrete application and context. Technically, the maximum size of a soft marker is only restricted by the size of the display, whereas the minimum size is restricted by the resolution respectively the pixel size of the display.

The structure and pattern of soft markers do not need to be fixed in advance, but can be dynamically assigned at runtime. In this way, the best subset of markers can be used for each individual setting and task. Besides optimizing the identification accuracy, this helps to avoid potential conflicts or allows to resolve them on the fly. In addition to optical markers, other useful information can be displayed.

In this sense, soft markers make the application of optical marker technology easy and transparent for the user. Users do not need to alter any hardware by carefully attaching markers at the right positions. Markers can simply be displayed by applications if and only when needed.

Although soft markers resolve some of the issues identified with static optical markers, they also introduce new ones. Soft markers are confined to surfaces with the ability to display and electronically alter at least a black and white image. This for example excludes objects without their own power supply and computation unit, but obviously includes many contemporary devices, such as TFT or CRT displays attached to workstations, laptops, PDAs, mobiles phones, electronic paper or consumer electronic devices possessing a sufficiently large pixel display.

In order to calculate the exact relative position of a marker, its size has to be known. This may lead to potential problems if soft markers are used across displays with different characteristics, or when markers are dynamically scaled. In these cases, the size of every marker displayed (and therefore the display’s pixel size) needs to be known. This information may directly be encoded in the marker pattern or provided over a separate communication link.

If patterns are dynamically assigned for the identification and differentiation of multiple objects, soft markers must be registered with some kind of infrastructure. This requires an active communication link and thus may require additional hard- or software components. For our display tracking application we assume that only displays with previously known characteristics are used and that these displays are controlled by devices networked to a ubiquitous infrastructure.

**Flashing Displays and Peeping Cameras: A two Step Approach**

The central idea is that displays wanting to know their position show a well defined set of soft markers which can be recognized and located by cameras in the environment. This calibration process is triggered whenever a display is added to or moved within the environment. This process may be started manually by the user, or automatically, i.e. whenever a new device with a display is connected the local network. Movements of devices may be observed by accelerometers, which are nowadays build into more and more mobile devices, especially laptops and PDAs.

In order to make our approach work reliably in larger areas, such as whole rooms, every possible display location must be observable from almost any potential direction by at least one camera. At the same time the maximum area covered by a single camera is limited by the size of the markers and the minimum image resolution required by the marker recognition system to work properly. Because of potential occlusion of markers by other objects or users, some redundancy in the coverage should be present. Obviously, when relying on fixed cameras a huge number of such cameras would be needed even for small rooms. Therefore we decided to use an array of remotely controllable pan/tilt/zoom network cameras to actively search the environment for markers.

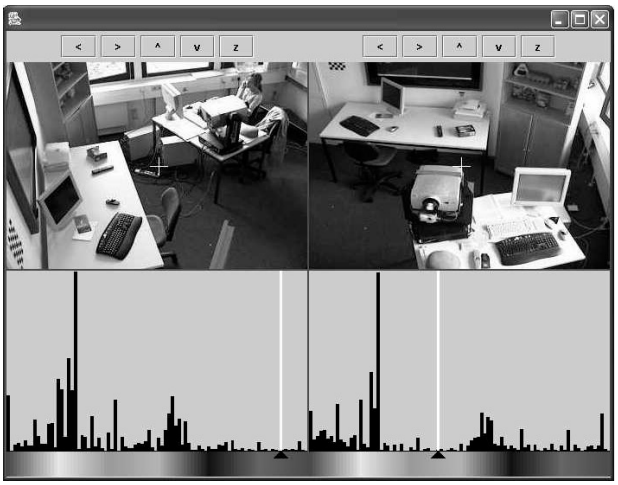

**Figure 3: Histogram analysis of camera images to find the optimal flashing color**

Nevertheless, scanning a potentially big room with only a small set of cameras for potentially small markers reminds of searching the famous needle in a haystack. For similar tasks humans have naturally developed a two step strategy which we adapt in our display tracking approach:

1. Instead of performing a detailed search right from the start, humans first look out for some easy to grasp characteristic structures that attract our attention. In our tracking application a display can grab the cameras' attention by making its whole surface blink. One can think of this blinking as a very big and simple soft marker being repeatedly turned on and off. Although the exact position and orientation of a display can not be computed from this blinking, it can be easily recognized by cameras from much greater distances than an actual marker. In addition, this blinking is very robust against partial occlusion, making it relatively easy to get at least a rough estimation of the searched display's position.

2. After having focused their attention on a promising candidate, humans further investigate the potential candidate and extract detailed information as needed. The same principle is used by our tracking approach after a blinking screen was found in the first step. The camera's zoom is used to "focus the attention" onto the identified screen. At this point, the blinking screen is changed to a set of markers as shown in figure 1. These markers, now hopefully seen by the camera in sufficient size, are used in the second step to determine the exact position and orientation of the display relative to the camera. Some redundancy is introduced by the display of multiple markers to account for potential partial occlusion caused by other objects or users.

As soon as the exact location of the display is detected in the second step, the soft markers are removed from the display and normal operation is continued.

## IMPLEMENTATION

In order to scan the room for flashing displays and detect markers on the displays found, we have equipped one room of our lab with several networked and remotely controllable cameras. One of them is shown in figure 2. Each camera provides an image stream with a resolution of 704x480 pixels. Furthermore, each camera's pan and tilt angle as well as its focal length can be controlled over a remote interface. For the recognition of soft markers displayed in the second phase of our approach we use the ARToolkit library (3).

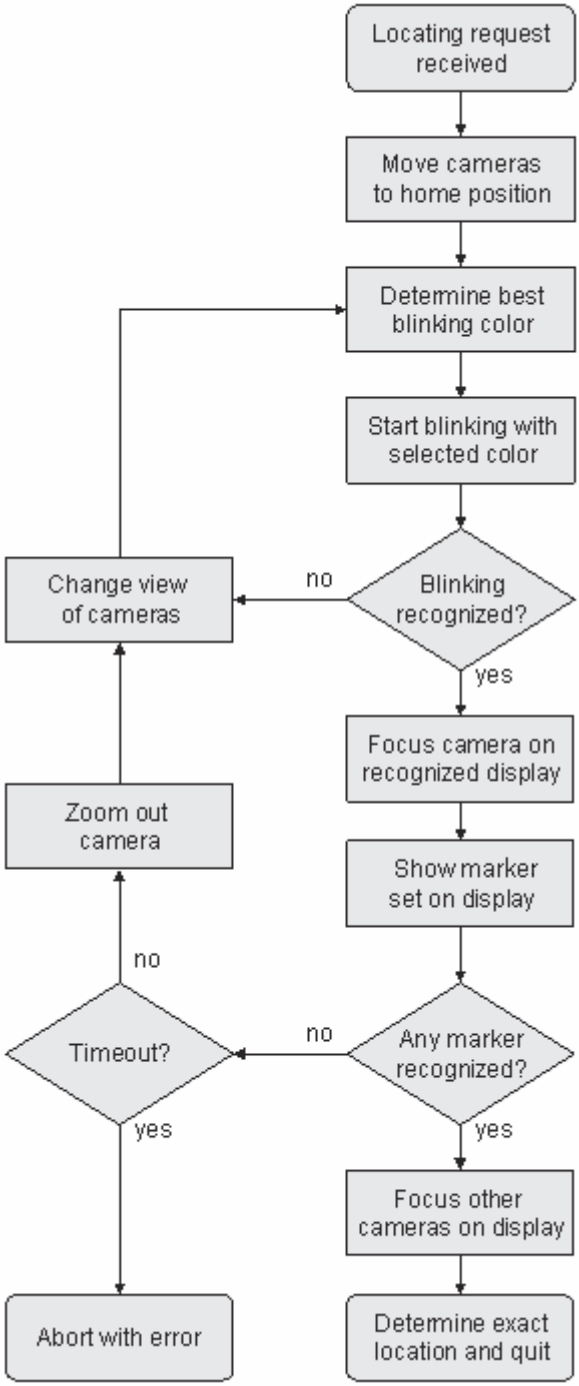


**Figure 4: Flow chart of two-phase recognition process**

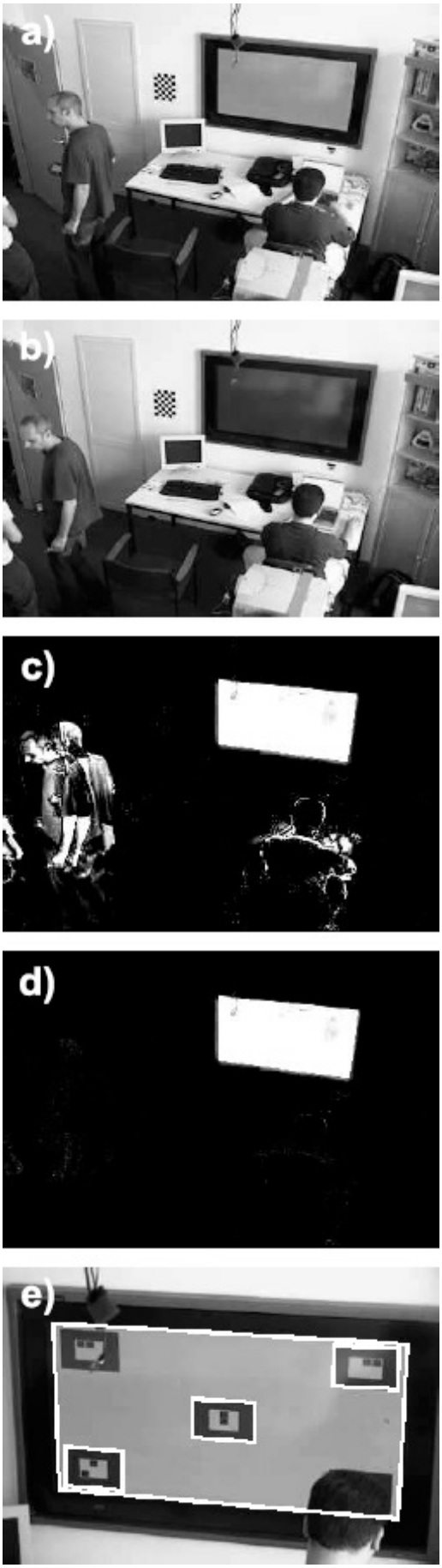


**Figure 5: Images as they are captured and processed during a sample run**

The central component of our implementation is the display locating service (DLS). All cameras in a room are connected to and controlled by this service. Displays can register with this service and issue a request to be located. Upon such a request, the DLS initiates and coordinates the locating procedure. The whole process is shown in figure 4 and explained in detail in the following paragraphs.

After the request has arrived at the DLS, the first phase is initialized. All cameras are brought into their home position; especially the zoom is set to telephoto. After this, the main loop of the first phase is entered: The DLS advises the display to start blinking with a certain color and frequency. Then all cameras are checked if they observe a blinking spot with the given color and frequency in the image. This is done by differential picture analysis. If yes, the tracking system continues with phase two, otherwise all cameras are newly oriented and the next scanning cycle is entered.

To improve recognition performance, we make use of the fact that the appearance of soft markers can be changed at runtime and thus can be optimized for the current situation. In our case, at the beginning of every iteration an optimal blinking color is determined. By causing the display to blink in a color rarely present in the rest of the environment, recognition performance can easily be increased by filtering the image for this particular color. This allows us to reduce environmental noise and movement. To determine the optimal blinking color, the histogram of the actual image taken is evaluated for each camera and new position. Figure 3 shows the output of a debug console for two cameras. In the upper half the original camera images are shown, while in the lower half the corresponding histograms are drawn. In each histogram, the optimal blinking color is marked with an arrow. By combining the feedback of each camera the DLS computes the optimal blinking color or a sequence of blinking colors if no single optimal color exists. This process is repeated for each iteration so that the blinking color may change multiple times during the first phase.

After at least one camera has observed the blinking screen and roughly estimated its size, phase two is entered. At the beginning of this phase, the DLS advises the display to stop blinking and display a pattern of five soft markers as shown in figure 1. At the same time the camera zooms in to maximize the size of the display in the image. In most cases the camera can now observe at least one of the five markers. With the size of the markers which has meanwhile been announced by the display, the rough position and orientation of the display can now be derived. If the camera doesn't recognize any markers, the original blinking is assumed to be a false alarm (caused for example by a reflection) and step one is reentered. This will repeat until the display is found or no good candidates are found in the blinking step.

Once the rough location of a display has been estimated by at least one camera, other cameras are also pointed at this position and are used to verify and improve the

recognition accuracy. After all calculations have been done, the display is notified by the DLS to remove the soft markers and display its original content again.

At the time of writing, not all parts of the approach described above are fully implemented. In the current state of our implementation the following restrictions hold: In the first phase we are currently applying the color filter only when searching for blinking displays. Knowledge about the blinking frequency is currently not used in the recognition process. Although this may be a slight disadvantage in theory, it turned out not to be a serious drawback in the vast majority of settings.

Although the zoom range of our cameras can nearly be continuously controlled, we currently use only four fixed focal lengths. This is due to the fact that for each focal length a separate calibration of the camera is needed to compute the relative positions of markers correctly. By restricting the focal length to four discrete values we reduced the calibration effort while preserving most of the locating performance. At the moment only the first camera recognizing a flashing display is used to look for markers in the second phase. This has proven to work well in most cases if a certain amount of inaccuracy is permissible and there is no object/user moving between phase one and two causing potentially severe occlusion.

## EXAMPLE

In this section we present an example run of our display localizing approach. A series of images taken and processed during the process is shown in figure 5.

In this example, the large plasma screen mounted on a wall in our lab requests to be localized. The images a) and b) are taken right after the optimal blinking color was discovered by the DLS and the plasma screen started blinking in that color.

From both images a differential image is generated. Image c) shows an contrast enhanced version of the differential image. As we can see clearly, the flashing display is visible together with some noise caused by other users moving around in the room.

To remove this noise, a color filter is applied to the input images a) and b) blocking all other colors than the blinking color of the display. The resulting differential image of the filtered video stream is shown in image d). Now the noise is almost gone, the rest of the noise is filtered out with some form of BLOB detection.

In the next step the bounding box of the flashing display is evaluated to estimate the rough size and position of the display. While the camera is focused and zoomed accordingly, the flashing on the display is stopped and replaced by the display of a set of soft markers redundantly denoting the four corners and the center of the screen.

Image e) shows the screen content as captured by the camera in the last step and evaluated by the use of the standard ARToolkit library to detect the size and location of the soft markers. From their position and orientation, the known display size, and the position and orientation of the camera, the exact location of the display can be deduced. The recognized soft markers and the inferred display bounds are highlighted in image e) by white lines. Through the presence of redundancy, partial occlusion caused by the wires in the upper left of the image or the user's head in the lower right can be compensated.

## APPLICATIONS

Knowing the position of displays in an instrumented environment is useful for many reasons. In this section we will shortly motivate the utility of a system like the one presented in this paper by discussing a general and one concrete application.

Generally, knowledge about the location of displays in an instrumented environment is central for every system utilizing the peephole metaphor described in Butz and Krueger (4). This metaphor is based on displays that act as "peepholes" into a virtual layer which ubiquitously spans the real world. This way, users can place and access arbitrary virtual items everywhere in the environment, as well as access information and services bound to real world locations and objects. Because a key point of this metaphor is the spatial correlation of real and virtual world, knowing the location of potential "peepholes" is important for every such system.

Another, concrete application of display tracking is the setup of an environment for the use of distant direct manipulation techniques to move digital objects between different devices. An example of such a technique is "Wiping" (Schneider and Butz, 12), which can be used to move items to a target system that is physically out of reach. To use this interaction technique, a wiping gesture is executed over the virtual items that should be moved. The intuition is, that the regarding items are accelerated by the wiping gesture and afterwards virtually "fly" through the environment, until they hit a potential target device (respectively screen). That way, the target can be specified by the gesture's speed and direction. It is obvious, that for such an interaction technique to work the locations of potential target displays need to be known.

## PROBLEMS AND FUTURE WORK

Our approach requires a line of sight between the camera and the displays. So far, we have provided a proof of concept for one camera, but we are aware that under realistic conditions displays may be occluded by objects or people in the environment. To overcome these problems we are currently extending the presented approach to work with multiple cameras.

The main idea is that in the first phase all cameras scan the room separately for unknown markers. Then, if a camera identifies the position of a possible marker, it can broadcast this information to all other cameras, which then in a second phase try to focus on the respective point in space to finally identify the marker. Important issues which need to be solved in this respect are camera calibration and synchronization. Also, each camera will have its own color histogram, which means that a compromise has to be found for the initial marker color. We plan to either use a compromise color that works for all cameras reasonably well or to realize a blinking pattern of distinct colors that are optimized for each of the different cameras.

Another interesting issue is the handling of movable displays, i.e. those of a tablet PC or PDA. One elegant way to approach this problem is to use accelerometers attached to or already build into the devices that locally detect whenever the device is moved. Markers can then be displayed while the device is under motion, which might even allow heuristics to track the displays permanently. The drawback of this approach is that the display cannot be used for other purposes while the device is moved, which i.e. means that feedback has to be provided by other means if necessary. Therefore, we also plan to look at a dual solution, where the marker is only displayed for a short time when the device comes to a halt after a movement.

Of further interest are optimal situations where the tracking may take place without disturbing users in the environment. Especially if the localization is not time critical, the scanning procedure can be postponed and carried out in the absence of users, or maybe even during the night, which might even improve the recognition of actively displayed markers.

## SUMMARY AND CONCLUSION

We have presented a robust two step approach for display tracking in instrumented environments. The method only uses displays and cameras present in the environment and doesn't require any specialized tracking hardware. It borrows its basic idea from a social protocol used by humans, which is designed to focus attention and preserve perceptive resources.

After describing the concept of soft markers, we have discussed our approach in detail and presented a prototypical implementation as well as a documented test run. We hope that this approach will provide a step towards unobtrusive, self-calibrating instrumented environments and thereby open the door a few inches further towards ubiquitous computing in our everyday environments.

## REFERENCES


1. Aoki H., and Matsushita S., 2000, "Balloon tag: (in)visible marker which tells who's who", Proceedings of Fourth International Symposium on Wearable Computers (ISCW00), 77–86

2. ART, May 2004, "Mars - mobile augmented reality system", http://www.ar-tracking.de/

3. Billinghurst M., Weghorst S., and Furness T., 1998, "Shared space: An augmented reality approach for computer supported collaborative work", Virtual Reality, 3(1):25–36

4. Butz, A., and Krueger A., 2003, "A generalized peephole metaphor for augmented reality and instrumented environments", In Proceedings of The International Workshop on Software Technology for Augmented Reality Systems (STARS)

5. Davison A., Mayol W., and Murra D., 2003, "Real-time localisation and mapping with wearable active vision", In The Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'02)

6. Harle R., and Hopper A., 2003, "Building world models by ray-tracing within ceiling-mounted positioning systems", In 5th International Conference Ubiquitous Computing (Ubicomp 03), 1–17

7. Intersense, May 2004, "Intersense is-900", http://www.isense.com/

8. Kishino Y., Tsukamoto M., and Sakane Y., and Nishio S., 2004, "A visual marker using computer displays for real space applications", In Advances in Pervasive Computing

9. Priyantha N., Chakraborty A., and Balakrishnan H., 2000, "The cricket location-support system", In Mobile Computing and Networking, 32–43

10. Rekimoto J., 1998, "A realtime object identification and registration method for augmented reality", in Asia Pacific Computer Human Interaction Conference (APCHI'98)

11. Rohs M., and Gfeller B., 2004, "Using camera-equipped mobile phones for interacting with real-world objects", in Advances in Pervasive Computing

12. Schneider M., and Butz A., 2005, “Wipe it! A direct manipulation technique for ubiquitous information items”, to appear in Proceedings of the IEE International Workshop on Intelligent Environments (IE 2005)

13. Wagner D., and Schmalstieg D., 2003, ”First steps towards handheld augmented reality”, in 7th International Symposium on Wearable Computers (ISWC’03)

14. Weiser M., 1991, ”The computer for the 21st century”, Scientific American, 3(265):94–104

15. Welch G., Bishop G., Vicci L., Brumback S., Keller K., and Colucci D., 1999, ”The hiball tracker: high-performance wide-area tracking for virtual and augmented environments”, in Proceedings of the ACM symposium on Virtual reality software and technology, 1–ff